HATNER: Nested Named Entitiy Recognition for German

Yulia Bobkova, Andreas Scholz, Tetiana Teplynska, Desislava Zhekova
CIS, Ludwig Maximilians University, Munich
{Yulia.Bobkova,Andreas.Scholz,Tetiana.Teplynska,D.Zhekova}
@campus.lmu.de

Abstract

This paper describes our classification and rule-based attempt at nested Named Entity Recognition for German. We explain how both approaches interact with each other and the resources we used to achieve our results. Finally, we evaluate the overall performance of our system which achieves an F-score of 52.65% on the development set and 52.11% on the final test set of the GermEval 2014 Shared Task.

1 Introduction

Named Entity Recognition (NER) is currently one of the most interesting and promising topics in NLP. It is commonly viewed as a subtask of information extraction (Nagy T. et al., 2011) and is a basis for many important applications, such as Coreference Resolution and Sentiment Analysis. NER by itself is no trivial task and NER for German is even more challenging, as the amount of available manually annotated data is limited. Additionally, capitalization is usually an important feature for detecting NEs. However, as nouns are generally capitalized in German, the usefulness of the capitalization feature is diminished. The quality of a NER system also strongly depends on its domain, as a system tailored to one specific domain generally performs worse on other domains (Poibeau and Kosseim, 2001). In this paper, we present a hybrid approach to NER in the implementation of HATNER.

Section 2 gives an overview of other approaches to NER. In section 3, we go into detail about the system requirements. In section 4, we give a short overview of HATNER and in Sections 4.1, 4.2 and 4.3 we go into more detail about the system. In section 5, we present our results and discuss them accordingly. Finally, in section 6 we conclude the our work.

2 Related Work

One of the earliest systems, which originally was intended for the English language only, is GATE (Cunningham et al., 2011). GATE itself is a conglomeration of different tools for NLP. One of these tools is ANNIE (a Nearly-New Information Extraction System) also described in (Cunningham et al., 2003). ANNIE uses finite-state algorithms and the JAPE language for regular expressions, as well as several gazetteers. During ongoing development support for more languages was added, amongst them German.

Another interesting approach and one of the best for English available today is the Stanford Named Entity Recognizer. It is based on a Conditional Random Field classifier and performs particularly well on the categories person, organization and location.

Lastly, specifically for German, there is one of the few freely available NER systems developed by Faruqui and Padó (2010). It is based on the previously mentioned Stanford NER and includes semantic generalization information from large untagged German corpora. It is one of the best NER systems for German available today.

Unfortunately, most state-of-the-art NER systems have not been developed with nested NEs in mind, which was newly initiated by the GermEval 2014 Named Entity Recognition Shared Task.

http://gate.ac.uk

http://gate.ac.uk/sale/tao/splitch6.html\#chap:annie

http://nlp.stanford.edu/software/CRF-NER.shtml

3 System Requirements

HATNER was specifically developed for the context of the GermEval 2014 Shared Task. The shared task specifies four main categories of entities to be recognized: person (PER), location (LOC), organization (ORG) and other (OTH), where OTH contains categories such as time, date, currency, religion and more. Each word or group of words in the data can qualify for any of these four categories, or none. For each of these four main categories, there also exists a part and derivative subcategory (labeled i.e. PERpart or PERderiv). Detailed information as to when a NE qualifies as part or derivative of a main category and the main categories themselves are specified by the NE annotation guidelines (Benikova et al., 2014). In short, one can define an entity as belonging to the part subcategory, if only a part of the NE belongs to a specific category, such as "Wembley-Tor", where Wembley is a LOC. The derivative category on the other hand mostly encompasses morphologically modified NEs, such as "Berliner" (as in: a citizen of Berlin, LOCderiv).
This results in a total of 12 possible categories for a NE. However, the aim of the shared task is not only to find NEs, but also to find NEs within said NEs. Hence, in a sentence like "Ich lese 'Das Tagebuch der Anne Frank'." there are two NEs: "Das Tagebuch der Anne Frank" (OTH), as well as "Anne Frank" (PER). Figure 1 shows an example of the annotation format as given in (Benikova et al., 2014). The second column depicts the word itself, followed by the NE tag for the first NE level and the NE tag for the nested NE level respectively. A tag starting with a B indicates the beginning of a NE. I indicates the inside of a NE and O the outside.

4 System Overview

Classification systems are generally more robust to change than rule-based systems and perform fairly well with an adequate feature set. However, they heavily rely on a large and qualitatively annotated training set. On the other hand, rule-based systems are very susceptible to changes and very time consuming to establish, but can better be tailored to specific needs. For these reasons,

#	http://de.wikipedia.org/wiki/Manfred_Korfmai		
1	Aufgrund	O	O
2	seiner	O	O
3	Initiative	O	O
4	fand	O	O
5	2001/2002	O	O
6	in	O	O
7	Stuttgart	B-LOC	O
8	,	O	O
9	Braunschweig	B-LOC	O
10	und	O	O
11	Bonn	B-LOC	O
12	eine	O	O
13	große	O	O
14	und	O	O
15	publizistisch	O	O
16	vielbeachtete	O	O
17	Troia-Ausstellung	B-LOCpart	O
18	statt	O	O
19	,	O	O
20	„	O	O
21	Troia	B-OTH	B-LOC
22	-	I-OTH	O
23	Traum	I-OTH	O
24	und	I-OTH	O
25	Wirklichkeit	I-OTH	O
26	"	O	O
27	.	O	O

Figure 1: Example of a tagged sentence in the final output file. (Benikova et al., 2014)

we propose a classification approach as the core of our system, which we also combine with a set of handcrafted rules specifically targeting the distinct NE types.

4.1 Preprocessing and Postprocessing

In order to provide our classifier with as many useful features as possible, we preprocessed each sentence. This included noun phrase identification, lemmatization and part of speech (POS) tagging. For this, we used the Python programming language as well as the NLTK toolkit and the TreeTagger (Schmid, 1994; Schmid, 1999).
As for postprocessing, the most important task is to ensure a well formed output file. Other than rules, a classifier is not guaranteed to always start a recognized NE with a beginning tag, but could instead start with an inside tag. Our postprocessing ensured the correct opening of each NE. We tried several different approaches, such

https://www.python.org
http://www.nltk.org
http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger

as conservative processing (converting I to O), neutral processing (I to B), optimistic processing (tag the previous word as beginning of the same category) or intelligent processing (considering noun phrases and sentence structures when deciding how to proceed). For the final tagging process we used conservative postprocessing as it provided the best results. Another step of postprocessing that we do is eliminating inner tags found by the second classifier which are not inside of any outer tag.

4.2 Classification

For the classification task, we use a maximum entropy classifier which is trained on the manually pre-tagged training set provided by GermEval.
We train two classifiers: one for the first NE level and the second one for the nested, NE level. In-between the classifier runs we perform a postprocessing step to ensure a well formed file for the second run. In order to achieve the best results, we devised and tested different features. The features of our final system are displayed in table 1.

For the second classifier, we use a subset of these features together with a feature which indicates whether an outer NE exists for the current token. The second run is also much more delicate. While the classifier is in fact encouraged to only tag tokens which were previously tagged as belonging to an outer NE, there is no guarantee for that. As we mention before, we compensate this with another post-processing step which handles inner tags which do not belong to an outer tag.

4.3 Rules

In the second part of HATNER, we specifically target areas the classifier had difficulties with, such as part and derivative forms of categories. With rules focusing on precision rather than recall, we intend to affect the results of the classifier as least as possible, while at the same time having a high confidence at actually improving or correcting a tag once all conditions of a rule had been met.
To keep the rules as specific as possible, it was not enough to use morphological and syntactic features only. We therefore created gazetteers for each of the four main categories. We extracted information from the German Wikipedia and also used the gazetteers available in the GATE system. Here, once again, German being the object of our studies turned out to be an added difficulty. Lists for the English language can easily be found, already available lists for German are scarce and inconsistent at best, non-existent at worst. Additionally, we need to detect which tokens may be part of a NE, so we lowercased the entries in the gazetteers, what led to the loss of information.

As for the gazetteers, we aimed at matching maximum length spans. However, during development, lists with less, but more specialised information performed better than large general lists. For example, after stripping down the names list to just common German and English names, we received much better results than with names from all over the world, as many of those tended to correlate with common, non-name words, in German.

Feature	**1^{st} Cl.**	**2^{nd} Cl.**
The token itself	yes	yes
The POS tag of the token	yes	yes
The POS tag of the previous token	yes	yes
The lemma of the token	yes	yes
Whether the token is within a NP	yes	no
The history of tags of the sentence	yes	yes
Outer NE tags assigned to this token	no	yes

Table 1: Feature sets of the first ($\mathbf{1}^{st}$ Cl.) and second ($\mathbf{2}^{st}$ Cl.) classifier.

5 Results and Evaluation

Table 2 shows the general results of the HATNER system, whereas table 3 shows the results of the part and deriv subcategories for each of the four main categories. We report results on the development set.

Setup	**Chunks**	**Prec.**	**Rec.**	**F1**
Classifier	Outer	71.26	44.98	**55.15**
	Inner	26.94	37.74	**31.43**
	Combined	64.59	44.44	**52.65**
Classifier + Rules	Outer	60.57	46.14	52.38
	Inner	19.12	30.66	23.55
	Combined	54.61	45.00	49.34

Table 2: General results of the system.

As can be seen in table 2, the final score of the classifier and rules combination is actually

http://de.wikipedia.org

Category	Classifier only	Classifier & Rules
LOCderiv		
outer	**75.43**	68.42
inner	**55.45**	19.69
LOCpart		
outer	29.51	**36.00**
inner	0.0	0.0
ORGderiv		
outer	0.0	0.0
inner	0.0	0.0
ORGpart		
outer	19.80	**55.63**
inner	0.0	0.0
OTHderiv		
outer	**46.15**	42.86
inner	0.0	0.0
OTHpart		
outer	10.53	**18.18**
inner	0.0	0.0
PERderiv		
outer	0.0	0.0
inner	0.0	0.0
PERpart		
outer	**10.53**	6.45
inner	0.0	0.0

Table 3: Subcategory results of the system.

performing worse than the classifier on its own. Interestingly enough, the classifier also performs better on nested NEs than the combined system. On the other hand, rules do improve some of the subcategories we actually designed them to improve. Table 3 shows that, while the derivative category seems to pose the most difficulties for either system, rules were able to compensate some of the weaknesses of the classifier in most of the part categories.

HATNER achieved 52.11% on the final test set based on the combined evaluation setting from table 2 (being M1, the official metric used by the task).

6 Conclusion

The paper presented the participation of our system at the GermEval 2014 Named Entity Recognition Shared Task for German. The results HATNER achieved on the development set indicate two facts: First, the combination of the classifier and the rules is worse than the classifier by itself. Second, rules are able to improve certain areas if tailored specifically to these areas. This leads us to believe, that, while this implementation of a combined system might have failed, it generally is possible and desirable. In our eyes, the key to achieving a combined system which actually performs better is to specialise rules even more. This would decrease the negative effect on the work of the classifier, while increasing the positive effects on the areas they would be designed to improve.

References

Darina Benikova, Chris Biemann, and Marc Reznicek. 2014. NoSta-D Named Entity Annotation for German: Guidelines and Dataset. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Cristian Ursu, and Marin Dimitrov. 2003. Developing Language Processing Components with GATE (a User Guide). http://gate.ac.uk, February.

Hamish Cunningham, Diana Maynard, and Kalina Bontcheva. 2011. *Text Processing with GATE*. Gateway Press CA.

Manaal Faruqui and Sebastian Padó. 2010. Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In *Proceedings of KONVENS 2010*, page 129. Semantic Approaches in Natural Language Processing.

István Nagy T., Gábor Berend, and Veronika Vincze. 2011. Noun Compound and Named Entity Recognition and their Usability in Keyphrase Extraction. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pages 162–169, Hissar, Bulgaria, September. RANLP 2011 Organising Committee.

Thierry Poibeau and Leila Kosseim. 2001. Proper Name Extraction from Non-Journalistic Texts. In *In Computational Linguistics in the Netherlands*, pages 144–157.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *International Conference on New Methods in Language Processing*, pages 44–49, Manchester, UK.

Helmut Schmid. 1999. Improvements in Part-of-Speech Tagging with an Application to German. In Susan Armstrong, Kenneth Church, Pierre Isabelle, Sandra Manzi, Evelyne Tzoukermann, and David Yarowsky, editors, *Natural Language Processing Using Very Large Corpora*, volume 11 of *Text, Speech and Language Processing*, pages 13–26. Kluwer Academic Publishers, Dordrecht.